DOCUMENT RESUME

ED 079 353 TM 002 952

AUTHOR	Dyer, Henry S.; And Others
TITLE	Feasibility Study of Educational Performance Indicators. Final Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	New York State Education Dept., Albany. Div. of Evaluation.
PUB DATE	Dec 67
NOTE	61p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Educational Research; *Evaluation Criteria; *Factor Analysis; Feasibility Studies; *Multiple Regression Analysis; *Performance Criteria; *Program Evaluation; Systems Analysis; Technical Reports

ABSTRACT

Progress made to date in efforts to lay the groundwork for establishing meaningful and on-going procedures for the assessment of the performance of educational systems in New York is reported. The need for educational performance indicators is that indicators would provide a series of measures to show how well educational systems are performing, that indicators would identify specific areas in which specific school systems need specific help, and that indicators could clarify what school are actually doing for and to different kinds of children. Four factors in measuring the performance of an educational system are input, educational process, surrounding conditions, and output. These factors constitute what is called the "student-change model" of an educational system. To reduce the number of indices, a factor analysis of the measures of student output characteristics should be carried out. Factor analyses would also be employed to reduce the number of measures of educational process, the measures of "hard-to-change" surrounding conditions, and the measures of "modifiable" surrounding conditions. The procedur proposed for deriving performance indicators for a given set of schools begins with a series of regression analyses involving as large a sample of school systems as appears feasible. A proposed pilot study to provide information crucial to the evaluation of underlying assumptions is discussed, and long-range plans are described. Appendix A to the report discusses factor analyses, and Appendix B concerns regression analyses and conversion of deviations to performance indices. (DB)

FILMED FROM BEST AVAILABLE COPY

ED 079353

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FEASIBILITY STUDY OF EDUCATIONAL PERFORMANCE INDICATORS

Final Report

to

New York State Education Department

from

Educational Testing Service

Henry S. Dyer
Robert L. Linn
Michael J. Patton

TM 002 952

Educational Testing Service
Princeton, New Jersey
December, 1967

Contents

		Page
I.	Preface	1
II.	Introduction	4
	The need for educational performance indicators Comparison of proposed method with usual approaches Background	
III.	Rationale	13
	The student-change model of an educational system Matrix of performance indices	
IV.	Procedure	20
	Factor analyses Regression analyses using school means Computation of educational performance indices based on means Within school regression analyses Alternative analyses	
V.	Utility of the Indicators	30
	Their function Hypothetical example	
VI.	Questions to Be Investigated by the Pilot Study	36
VII.	Recommendations	40
	The general plan Phase one--study involving QMP data Phase two--study involving PEP and BEDS data Implementation	
VIII.	Long Range Plans	51
	Developing an operational system Continuing modification and simplification Prospect for cost-benefit analyses Experimental inquiry	
Appendix A:	Factor Analyses	55
Appendix B:	Regression Analyses and Conversion of Deviations to Performance Indices	58

I. PREFACE

In 1967 Educational Testing Service was asked by members of the New York State Education Department to explore the feasibility of developing a coherent and useful means of assessing the performance of school systems in the State of New York.

Numerous visits for this purpose were made to the State Education Department in Albany by ETS staff members from February through August 1967. During the meetings that took place, matters of policy and management pertaining to the development of a system of educational performance indicators were considered. On April 14, 1967, Memorandum #1 on "The Conception and Functions of Educational Performance Indicators" was submitted to the Department for consideration, and on April 30, 1967, Progress Report #1 was completed.

These two documents, together with the advice and assistance of various members of the Department staff, formed the context in which the idea of statewide indices of the performance of school systems was evaluated. In addition, data from the on-going research and evaluation programs of the Department were investigated for their potential inclusion in a proposed pilot study that would demonstrate the feasibility of educational performance indicators. Department personnel concerned with the Quality Measurement Project (QMP), the Pupil Evaluation Program (PEP), and the Basic Educational Data System (BEDS), provided invaluable assistance in identifying data sources and suggesting and commenting upon the proposed operating procedures that were presented in Memorandum #3, a "Tentative Plan for a Pilot Study of Performance Indicators" as contained in Progress Report #2.

Through the arrangements made by the Department, a group of consultants was brought to Albany to review these memoranda and to discuss their reactions with ETS and Department staff members. The following consultants were able to attend the meeting: Mrs. Winthrop Davenport, Dr. Noble Gividen, Dr. Jack Merwin (representing Dr. Ralph Tyler), and Dr. Alexander Mood. The consultants who provided written critiques of Memorandum #1 included: Dr. James S. Coleman, Mrs. Winthrop Davenport, Dr. Neal Gross, Dr. Alexander Mood, and Dr. Seymour Wolfbein.

Following the meeting with the consultants, Dr. Alan Robertson, Director of Evaluation, in the State Education Department, was asked to coordinate the proposed pilot demonstration study. A group of staff members in the Department was also appointed as an advisory committee to assist Dr. Robertson.

On July 14, 1967, an ETS staff member met in Albany with Dr. Robertson and Dr. Anderson to discuss preparation of those data from the QMP and PEP projects that would be used in the pilot study. During that meeting it was suggested that prior QMP research findings might well answer some of the questions raised in Memorandum #3. On August 11, 1967, a second ETS representative discussed these possibilities with Drs. Robertson, Armstrong, and Wohlferd in Albany.

To date, substantial progress has been made in clarifying the rationale for the development of educational performance indicators and in establishing procedures for conducting a pilot study in the State Education Department that would provide an empirical demonstration of their feasibility. The continuing dialogue between the consultants from ETS and the members of the State Education Department has been instrumental in accomplishing the tasks involved in such an endeavor.

This report will detail in summary form the progress to date of both Department of Education and ETS representatives in their efforts to lay the groudwork for establishing meaningful and on-going procedures for the assessment of the performance of educational systems in New York. The major recommendation made in the report urges that the next step in developing an operational system of educational performance indicators must be a large scale pilot study. Such a study would provide a kind of proving ground on which to test the notions contained in this report.

II. INTRODUCTION

The Need for Educational Performance Indicators

All too frequently attempts to assess the performance of an educational system fail to make explicit how observed or inferred educational outcomes--desirable or otherwise--are related to both the characteristics of students and the educational experiences and conditions assumed to be antecedent to the outcomes. In the absence of a useful logic of relationship between the attributes of the human beings who influence and are influenced by the educational system in which they participate and the system itself, our ability to make sound decisions to improve the educational progress of students is attenuated. At least three reasons can be stated in support of developing a systematic approach to assessing the performance of educational systems that would alleviate this state of affairs.

The first reason is simply that indicators would provide a series of measures to show those responsible for educational systems how well their systems are performing. These measures would be aimed at identifying and highlighting the points where a school system is falling short in meeting the developmental needs of its pupils. They would constitute measures that take due cognizance insofar as possible of the conditions in which a school system must operate. That is, they would be "fair" in the sense that the performance of any school system would be compared only with other systems similar to itself. One important purpose, then, of the educational performance indicators is to obviate the mis-use of test data and other performance measures that might lead to ill-conceived and invidious comparisons, as is so often the case under present circumstances.

The second reason is that indicators would identify the specific areas in which specific school systems need specific help. Such help could take the form of professional services. It could also take the form of State funds earmarked for specific educational improvements.

The third reason for the indicators is that they could clarify in concrete and highly visible ways what schools are actually doing for and to different kinds of children so as to raise forcibly in the public mind--i.e., in the minds of the policy-makers that sit on committees of school faculties, school boards, legislative assemblies, PTA's, and the like--what specific goals they want their schools to reach and how much they are willing to pay in dollars and in sacrificed opportunities for the accomplishment of these goals.

Comparison of Proposed Method with Usual Approaches

It will be helpful to consider, now, some of the ways in which educational systems are often conceptualized and assessed in order to highlight some of the differences between customary approaches and the approach advocated in this report.

The typical method of sizing up an educational system rests on a miscellaneous collection of unsupported, unarticulated, and often unconscious hunches. It is essentially a seat-of-the-pants approach which operates by assuming an educational organization to be in good working order:

--IF the school plant is in good shape

--IF the most up-to-date equipment has been installed

--IF there are enough textbooks to go around

--IF the teachers meet certification requirements

--IF the program includes all the approved courses and special services

--IF the pupil-teacher ratio is no greater than 25:1

--IF the library is well stocked

--IF teachers' salaries are above the national average

--IF the day-to-day administration of the system is divorced from local politics

and so on through a long string of additional and highly problematical IF's.

What is wrong with this approach to measuring the performance of educational systems? Three things mainly.

Mistaking means for ends. Probably its worst defect is that it traps too many people, including professional educators, into mistaking means for ends. It rivets their attention exclusively on the instrumentalities of the system as though the instrumentalities were ends in themselves. It may not even raise in their minds the question whether the instrumentalities -- the books, the buildings, the teachers, the language labs, the fancy new curricula, etc. -- are helping or hindering or having no impact at all on the intellectual, social, and personal development of students. The efficiency of the system is measured in terms of how much gadgetry the educational dollar is buying rather than how much change in pupils the educational process is producing.

Dependence on unchecked assumptions about means. There are of course many educators and educational policy-makers who are capable of rising above the kind of sterile thinking that equates the purpose of education with getting more and cleaner toilets or raising teachers' salaries or installing teaching machines and computers. These people concede that buildings and books and teachers are educational means, not ends; that we do not run schools to make jobs for teachers; or to make profits

for builders, bus operators, textbook publishers, and computer manufacturers; or even to satisfy the intellectual compulsions of curriculum innovators and the inventors of new administrative arrangements like team teaching and flexible scheduling and the extended academic year. Educational systems, they admit, are supposed to be run for the benefit, not of the educators, but of the people to be educated. We must therefore assume, they say, that teachers with MA's are more helpful to pupils than teachers without MA's; that clean, bright classrooms are conducive to clean, bright minds; that better-organized courses will produce better organized citizens, and so on.

This type of thinking does not confuse means with ends; it simply imagines, on the basis of intuition uninhibited by data, that certain causative connections between ends and means must exist, even though the relationships between the two have never been explicitly or adequately examined. It says that, in the absence of empirical evidence, it is reasonable to assume that, for instance, better-trained teachers will make for better-educated pupils, that a foreign language program beginning in the third grade will teach more children more French or Spanish or Russian than one that begins in the ninth grade, or that singing the folk songs of different ethnic groups will help children better understand and appreciate children different from themselves in ethnic origin.

The trouble with such "reasonable assumptions" is that, at worst, they may be wrong; or, at best, they may be right in some circumstances but wrong in others. The teaching of foreign language in elementary school may induce in some children a life-long horror of foreign language study. Some teachers who concentrate on the piling up of credentials by means of graduate study may become less interested

in the welfare of their pupils and more interested in climbing up the salary scale.

In short, "reasonable assumptions" about the relationships between means and ends in education can turn out, upon examination, to be quite unreasonable. As early as 1897, Joseph Mayer Rice demonstrated how unreasonable reasonable assumptions can be in his famous article, "The Futility of the Spelling Grind." Up to then educators assumed, quite reasonably, that the more time and effort a teacher put on spelling, the better her pupils would be able to spell. Rice showed that this was just not so.

Dependence on unchecked assumptions about ends. Unreasonable assumptions can also work in reverse. For example, if the students from System A tend to score lower on reading tests than the students from System B, it is often assumed that the teaching of reading in System A is less effective than the teaching of reading in System B. Or if the incidence of juvenile delinquency in System X is greater than the incidence of juvenile delinquency in System Y, it is assumed that System Y is doing a better job of character training and inculcating the attitudes of good citizenship than is System X. Such assumptions can be wholly unreasonable. Looking solely at what pupils are like as they emerge from any phase of an educational system tells nothing whatever about how the system is functioning. One has to know in addition what relationships may exist between the characteristics of youngsters as they come out of any phase of the system and the characteristics with which they entered that phase of the system. One also has to know with considerable specificity what went on inside the system that might have brought about any changes in those characteristics and what went on outside the system

that might have facilitated or impeded any such change. This requires exact knowledge of the social setting in which educational events occur.

In summary, then, there are three main kinds of fallacy that customarily crop up in the assessment of educational systems: (1) the kind that confuses means with ends (2) the kind that makes "reasonable assumptions" about causative connections between means and ends without ever checking the reasonableness of such assumptions (3) the kind that assumes that knowledge of how students perform as they emerge from any phase of an educational system is a sufficient basis for assessing the effectiveness of the system.

Obviating the fallacies by means of performance indicators. The idea behind the development of educational performance indicators is to achieve a method of measuring the performance of educational systems in a way that obviates these fallacies. There are two main ideas that underlie the indicators: first, that a measure of system performance must be made up of measured changes in the students the system is supposed to serve; second, that these system measures must be such as to permit reasonable but tentative inferences about possible relationships between the changes that occur in students and the attributes of the social settings (e.g., school, home and community) under which they occur. The indicators cannot purport to identify unequivocally the cause of change; they should however suggest hypotheses that might be explored and what steps might be taken to increase the effectiveness of educational systems. Thus, educational performance indicators are not to be thought of simply as interesting numbers, but as an indispensable information base for planning programs to improve the schools.

Background.

There are at least four previous studies which seem particularly relevant to the task of developing an operating system of educational performance indicators. The first of these was conducted by Samuel M. Goodman under the

auspices of the New York State Education Department.[1] This, of course, is the well-known Quality Measurement Project.

This study is based on IQ and achievement test data for some 70,000 students in grades 4, 7 and 10 in 103 school systems. Data for grade 7 were used for most of Goodman's correlational analyses. Goodman found a substantial relationship between socioeconomic status and achievement (a correlation of .61). He also found teacher experience and per pupil expenditure to have substantial relationships with achievement (correlations of .56 and .51 respectively). Of greater significance, however, is the fact that even after the presumed effect of socioeconomic status was controlled statistically, the relationships of teacher experience and per pupil expenditures were still strong enough to make plausible the hypothesis that teacher experience and per pupil expenditure have something to do with how much children learn in school (partial correlations of .37 and .31 respectively). Data such as these enable one to start with something besides "unchecked assumptions about means."

The Quality Measurement Project resulted in a wealth of data, only a small portion of which could be analyzed by the Goodman study. During the last 10 years, these data have generated many analyses and reports both of a formal and informal nature, some of which involve longitudinal data which were not available for Goodman's analyses. The results of these analyses should provide answers to a number of basic questions about the procedures which are developed in this report. As is recommended below, these existing analyses should constitute the first line of attack on the questions that need to be investigated in the development of an operational system of educational performance indicators.

---

[1] Samuel M. Goodman, The Assessment of School Quality. Albany: The New York State Education Department, March, 1959.

The second study which is considered particularly pertinent for our concerns is considerably more recent. The recent results of a longitudinal study of 658 Project TALENT students conducted by Marion F. Shaycoft provide evidence that students in some schools show greater increases in performance than students in other schools even when prior performance is taken into account.[2] The extent to which these differences can be attributed to differences in the quality of the schools per se and to differences in the conditions of the communities surrounding the schools remains problematic. Miss Shaycoft argues quite convincingly, however, that it is likely that much of the variation in student achievement between different schools can be attributed to school characteristics which themselves differ from school to school.

Any discussion of studies that have investigated differences among schools in terms of student achievement must include a consideration of the Coleman survey.[3] One of the salient conclusions of this study is that the differential performance of students in different schools "appears to arise not principally from factors that the school system controls, but from factors outside the school proper."[4] This conclusion is based almost exlusively on a measure of verbal ability which is known to be highly associated with a student's home background. It seems reasonable to expect that the differential influence of schools would be more marked on othermeasures such as mathematics or literature, as was indeed the case in Shaycoft's results. It also should be noted that Coleman's conclusion is based

---

[2]Marion F. Shaycoft, The High School Years: Growth in Cognitive Skills. Pittsburgh: American Institutes for Research and School of Education, University of Pittsburgh, 1967.

[3]James S. Coleman, et al. Equality of Educational Opportunity. Washington, D.C.: U. S. Office of Education, 1966.

[4]Ibid., p. 312.

on analyses within eight relatively homogeneous groups that might obscure some relationships between school factors and achievement.

The final study to be considered here is commonly known as the Pennsylvania Project which was conducted by ETS.[5] The logical framework for and approach to evaluation of schools that was developed in the Pennsylvania Project are essentially the same as those to be developed in the present report. Three fundamental tenets of the Pennsylvania Project which are taken as the starting position of the present report are: (1) the quality of a school must be evaluated in terms of student performance, (2) to compare two or more schools in terms of student performance, adjustments must be made to take into account differences in the performance of the same students at an early point in time and differences in the hard-to-change conditions of the surrounding community and (3) when comparing a set of schools that, on the one hand, are demonstrated to have equivalent advantages and/or handicaps but, on the other, show differences in level of student performance, clues about how the less effective schools might improve their situation can be gained by observing where these schools differ from their more effective counterparts with regard to certain modifiable surrounding conditions and educational processes.

---

[5]Educational Testing Service. A Plan for Evaluating the Quality of Educational Programs in Pennsylvania: A Report from Educational Testing Service to the State Board of Education. Princeton, N. J.: Educational Testing Service, 1965.

III. RATIONALE

The Student-Change Model of an Educational System

For the purpose of measuring the performance of an educational system, we may conceive of four factors in its operations. These four factors are input, educational process, surrounding conditions, and output. They constitute what we are calling the "student-change model" of an educational system. The inter-relationships of the factors are suggested by the chart in Figure 1. The four factors are defined as follows:

Input. The input of the system consists of all the characteristics of pupils as they enter any particular phase of an educational program; their mastery of the basic cognitive skills, their health and physical make-up, their knowledge, their attitudes, interests, social behavior, aspirations, etc. We do not distinguish between inherited characteristics and those that students have acquired from the impact of the environment, since the distinction is problematic, confusing, and irrelevant for the present purpose. Input consists of descriptive measures of students as they are at a given point in time. These descriptive measures involve no assumptions whatever about how the students got that way.

Output. The output of the system consists of all the measured characteristics of the same pupils as they finish any particular phase of an educational program. Again, we are concerned with all the cognitive, noncognitive, and physical characteristics of the pupils, whether these are to be attributed to experiences in school or elsewhere. That is, we do not confine our attention to those output characteristics that we assume might have been or could have been affected by the school experience, for our basic concern is to attempt to sort out the changes in pupils (good and bad) which might reasonably be attributed to the events in the educational setting as such from those that might be attributed to the non-school environment.

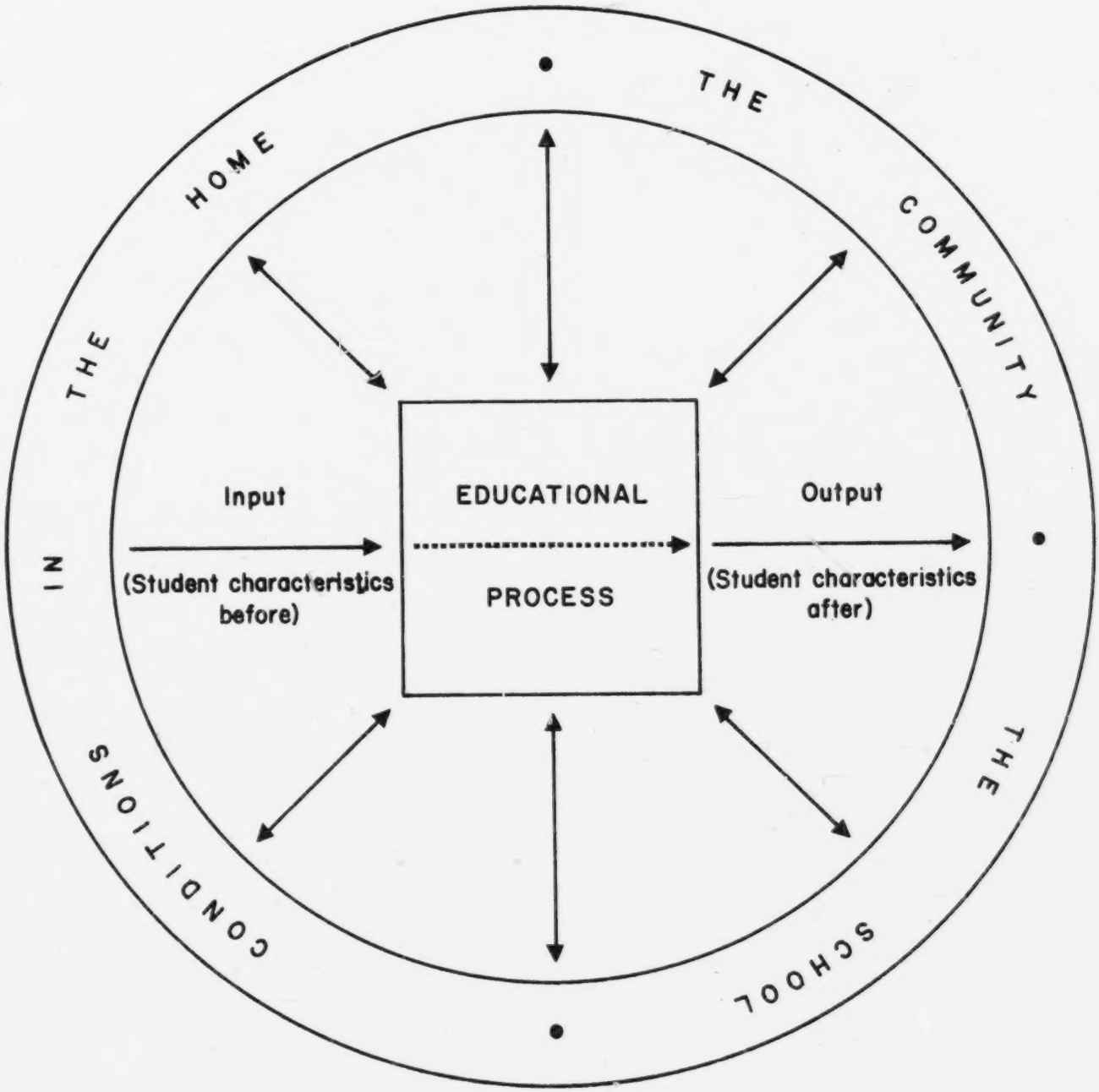

Figure 1

Factors in the Student-Change Model
of an Educational System

Educational Process. Educational process consists of all the activities in a school setting which are intended explicitly to bring about changes in the pupils. Lessons in arithmetic, organized and informal athletics, educational and vocational counseling, independent study, homework, participation in student government, the health program, viewing of film strips, tests and examinations, the marking system, conferences with parents -- all of these and many more are observable events in an educational setting. It must be kept in mind that the effects of the educational process in a school district, or in a single school, or even in a single classroom are not necessarily, or likely to be, uniform for all pupils. Any observations or measures of these events must take account of two major variables beyond the description of the events themselves: (a) the variety of the teaching-learning activities aimed at furthering pupil development (b) the amount of differentiation in these activities from pupil to pupil. Or to put the matter another way, educational process is to be characterized not only by what goes on in a school system on behalf of the pupils, but also by the richness of the program and by the effort to adapt the components of the program to the developmental needs of each individual student.

Surrounding conditions. The surrounding conditions are all of those influences in the educational environment that are likely to affect for better or worse how and what teachers teach, and how and what pupils learn. They are of three kinds: home conditions, school conditions, community conditions.

Home conditions include such matters as the level of education of the pupil's parents, the level of the family income, the size of the family, the degree to which the family understands and values education, the actual physical condition of the house where the child lives, the quality of treatment the child receives from parents and siblings, etc.

School conditions include both ecological factors like the school building, the number of pupils and teachers, the equipment with which they have to work, the spaciousness of the classrooms, and psycho-social variables like the training, experience, and attitudes that teachers bring to their work, and the general atmosphere -- the values and customs -- that pervades the school.

The distinction between school conditions and what we are calling events of the educational process is sometimes a fine one. We think, however, it is a distinction worth making wherever possible. A well-stocked school library has to do with observable school conditions; the use to which the library is put has to do with observable features of the educational process. An English teacher's love of literature (i.e., her attitude) is an inferred condition of pupil learning; the manner in which she tries to impart her love of literature is an observable feature of the educational process.

Community conditions include such things as the size of the community, the amount of taxable wealth available for support of the schools, the degree to which the citizens are willing to support the schools, the density of the population, the number and quality of social agencies, the presence or absence of industry, the employment rate, the crime rate, etc.

A dimension running through all the conditions that surround the educational process is the degree of their modifiability. Are they easy or hard to change? As we shall show later, this distinction is important in the development of educational performance indicators.

Matrix of Performance Indices

Given the student-change model of an educational system as described above, any measures of the performance of such a system necessarily will be complex.

They are complex technically because the system itself is complex. Any adequate measure of the performance of an educational system must simultaneously take account of input and condition variables as well as output variables. Furthermore, no single measure of system performance can usefully characterize the system. To depict how a system is performing we require a matrix of performance indices. One dimension of the matrix is a time dimension expressed in phases of the educational system or years of schooling, e.g., pre-primary, primary, intermediate, secondary, or years 1-3, 3-5, 5-7, etc. The other dimension of the matrix consists of categories of student characteristics, e.g., cognitive development, attitudinal development, interpersonal behavior, etc.

Categories of performance indices. The categories of performance indices are indispensable as bases for defining educationally meaningful goals or objectives of a system though they themselves are not the goals. It is not until one has decided what is to be changed and the directions such changes are to take that one has defined the goals of the system.

An illustrative matrix is shown schematically in Figure 2. The illustration suggests six major categories of performance indices. These might well be expanded to 20 or 30 subcategories, or possibly reduced to two or three more global categories depending on the dimensionality of the pupil characteristics observed.[6] The main point is that in characterizing an educational system it is of the utmost importance to measure insofar as is possible how all the characteristics of the pupils change as they go through the system.

---

[6] By implication, the procedures call for factor analyses of student characteristics as well as of the conditions under which they occur.

Phases of education by years in school	Major categories of performance indices*					
	A Physical	B Cognitive	C Social	D Moral	E Vocational	F Recreational
Year 11 to year 13	$PI_{A.13}$	$PI_{B.13}$	$PI_{C.13}$	$PI_{D.13}$	$PI_{E.13}$	$PI_{F.13}$
Year 9 to year 11	$PI_{A.11}$	$PI_{B.11}$	$PI_{C.11}$	$PI_{D.11}$	$PI_{E.11}$	$PI_{F.11}$
Year 7 to year 9	$PI_{A.9}$	$PI_{B.9}$	$PI_{C.9}$	$PI_{D.9}$	$PI_{E.9}$	$PI_{F.9}$
Year 5 to year 7	$PI_{A.7}$	$PI_{B.7}$	$PI_{C.7}$	$PI_{D.7}$	$PI_{E.7}$	$PI_{F.7}$
Year 3 to year 5	$PI_{A.5}$	$PI_{B.5}$	$PI_{C.5}$	$PI_{D.5}$	$PI_{E.5}$	$PI_{F.5}$
Year 1 to year 3	$PI_{A.3}$	$PI_{B.3}$	$PI_{C.3}$	$PI_{D.3}$	$PI_{E.3}$	$PI_{F.3}$
Birth to year 1	$PI_{A.1}$	$PI_{B.1}$	$PI_{C.1}$	$PI_{D.1}$	$PI_{E.1}$	$PI_{F.1}$

*Interpretation of cell entries: $PI_{B.5}$ is the performance index for Area B (i.e., output representing change in cognitive skills or function) for the two-year phase ending at year 5.

Figure 2

Illustrative Matrix of System Performance Indices

Phases in the Educational Process. Similarly, the illustrative matrix suggests seven phases in the educational process. One could expand the number of phases so as to take account of shorter intervals in the pupil's career, say six months; or one could reduce the number by lengthening the intervals between observations. The number of phases to be accounted for is largely a matter of practicality and of the length of time required to generate increments of change that can be measured with reasonable reliability. Again, however, the main point is that since it is useful to think of an educational organization as a dynamic system that operates at several levels at once, its performance must be measured over time at each of the several levels in order to provide usable management information for the continual improvement of the system and for keeping it up-to-date with the developing needs of society and of the pupils.

IV. PROCEDURE

Factor Analyses

On the one hand, it is important to measure as many as are feasible of the pupil characteristics which the school might be expected to change. On the other hand, if the educational performance indices are to have much practical utility there must be a reasonable limit to their number. To reduce the number of indices a factor analysis of the measures of student output characteristics should be carried out. The factor analysis would permit the selection of a reduced set of output measures for which performance indices would be developed. A more detailed and technical description of the factor analysis procedure to be used and the procedure for selecting a reduced set of variables is presented in Appendix A.

Factor analyses would also be employed to reduce the number of measures of educational process, the measures of "hard-to-change" surrounding conditions, and the measures of "modifiable" surrounding conditions. The available measures would first be placed into one of the three categories above, and a separate factor analysis performed for each category. The results would then be used to reduce each set of variables by (a) combining measures and (b) possibly dropping some measures. The same factor analysis procedure would be used for the measures of student output, except for the method of reducing the number of variables (see Appendix A for technical discussion).

Regression Analyses Using School Means

The procedure we propose for deriving performance indicators for a given set of schools begins with a series of regression analyses involving as large a sample of school systems as seems feasible. For a 6 x 7 matrix, like the one in the illustration, we would require 42 regression analyses -- one for each cell in the matrix.

For each regression analysis, the dependent variable would be the mean of a given output (say, scores on a reading test) taken at the close of a particular educational phase, e.g., years 7-9. The independent variables would be of two kinds: averages of all the pupil input characteristics taken at the beginning of the phase and measures of all the hard-to-change surrounding conditions that have obtained during the period of time covered by the phase. The measures of both dependent and independent variables would be school system means. (A more detailed technical description of this phase of the analyses is presented in Appendix B.)

The purpose of the regression analysis is to obtain a best weighted composite of all input and condition factors in order to best predict school system outputs given the various combinations of advantages and disadvantages. The performance index for any system is a number showing how its actual output compares to the output predicted from all the input and condition factors.

Computation of Educational Performance Indices Based on Means

Figure 3 gives an illustration of how the performance indices would be assigned for any particular student output characteristic. In this example the actual output is the set of school system means of sixth graders on a reading test. The predicted output is based on the best combination of the input (which consists of system means for fourth graders on a reading test and a number of other pupil characteristics) and the condition factors.

The performance indices that are assigned to school systems are determined by the sections of the scatter plot formed by the diagonal lines. If a school system (represented by a dot) falls below the lowest diagonal line, it is assigned a performance index (PI) of 1; if the school system is between the lowest and next lowest line, it is assigned a PI of 2, and so forth. The diagonal lines which are

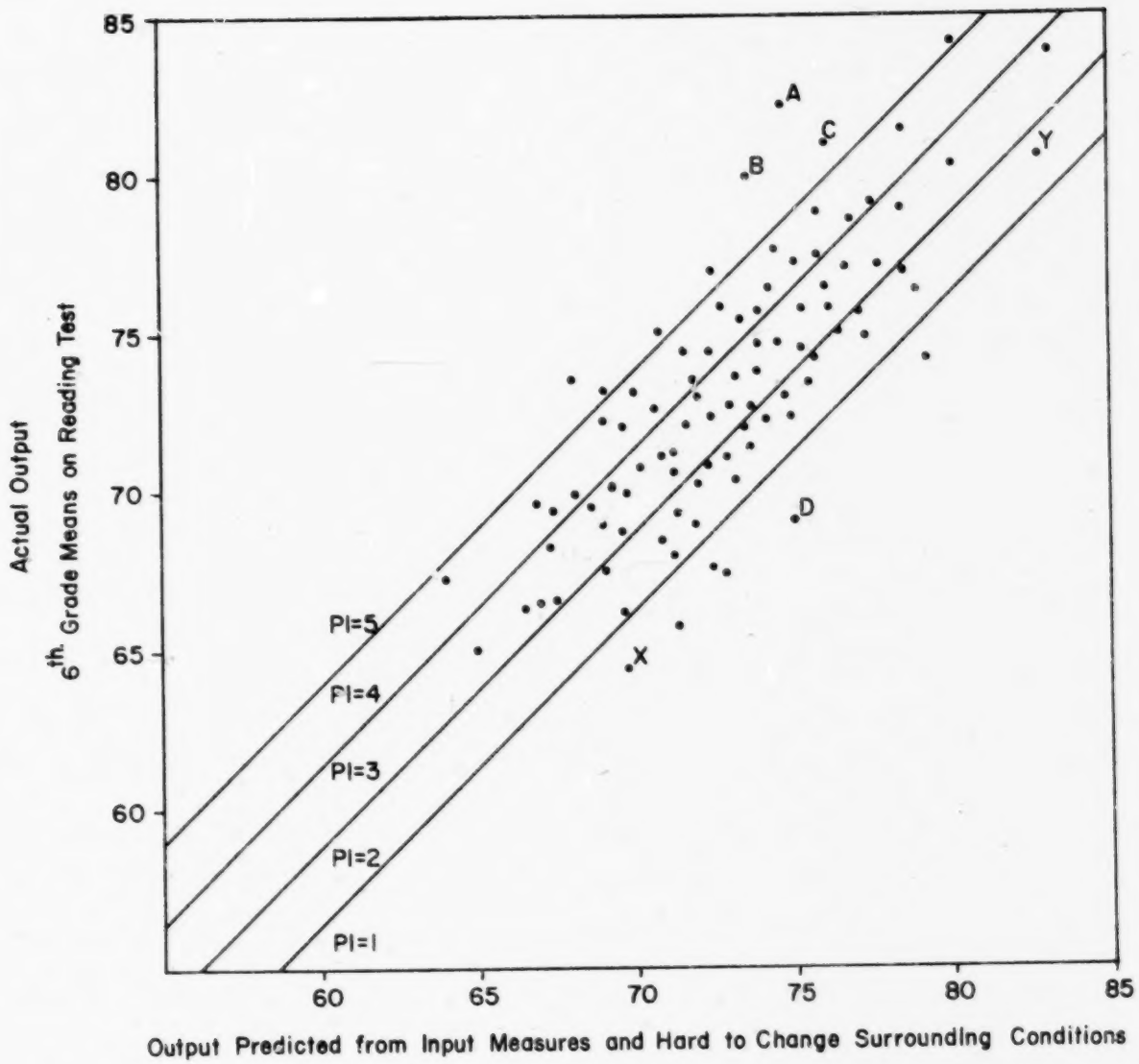

Figure 3

Illustration of the Assignment of Performance Indices to 91 School Systems

used to create the several bands and thereby to assign the performance indices to the schools are parallel to the regression line used for obtaining predicted outputs. The distance of each of these lines from the regression line is determined by the estimated accuracy of the output for an average school system. (The actual computational procedure is specified in greater detail in Appendix B.)

One of the main ideas behind this notion of performance indicators is that the top systems set tentative standards for other systems with approximately equal predicted outputs in any particular performance area for any specified phase of education. For example, with respect to quality of instruction in reading during the period from Grade 4 to Grade 6, Systems A, B, and C in Figure 3 might, with some reason, be regarded as pace-setters for other systems with approximately equal predicted output (say, in the range of about 70 to 75). Thus it would seem reasonable for school system D which has a predicted Grade 6 reading output similar to that of A, B, and C, to look at these schools for clues about ways that it might seek to improve the reading of its pupils. It probably would not be of much value, however, to compare school systems X and Y to A, B, and C, for this purpose since X and Y are dealing with quite different input and presumably quite different condition factors. Comparison of schools with approximately equal predicted outputs is a kind of "rough justice," but it is far and away superior to the kind of blind approaches currently used to appraise school systems, -- approaches which rely on such unadjusted figures as the number of students winning college scholarships or the number exceeding the norm on national testing programs.

An extremely important feature of the educational performance indicators, as we see them, is that they are based on relative <u>gains</u> in specific types of measured student performance during specified periods of schooling. According to this conceptualization of the indicators, the PI assigned to a school for the

quality of its reading instruction during Grades 4 to 6 could be derived from its position relative to one group of schools working under similar input and conditions for that period, but the PI assigned to the same school for the quality of its reading instruction during Grades 7-9 could be derived from its position relative to quite a different group of schools having comparable inputs and conditions during that later period. By the same token, the PI assigned to a given school for its Grades 4-6 performance in reading could be based on its position with respect to one set of schools, but the PI assigned to the same school for its Grades 4-6 performance in health education could be based on its position with respect to a different set of schools.

It must be borne in mind that, in spite of all the refinements one may build into the procedure for deriving a performance indicator for a school or school system for measuring any part of its program at any level, the index can never be a perfectly reliable and valid measure of system performance. It is not perfectly reliable because the differentiation among systems is in part at least the result of random error in the means. It is not perfectly valid because some important independent variables may have been overlooked in the regression analysis, with the result that some systems would have higher or lower predicted outputs if these variables were included. It is conceivable, for example, that System A's "true" predicted reading output should be about the same as that of System Y, in which case its "true" performance index for reading would be a "2" rather than a "5."

This emphasizes the fact that the educational performance indicators obtained for any educational system must be interpreted -- as any such measures must always be interpreted -- with due caution. They should be regarded strictly as indicators -- i.e., as pointers or clues -- for identifying systems most likely to be off the standard of predicted performance in some category and therefore most likely to be in need of help from the State, both professionally and financially.

Within School Regression Analyses

The indices based on school means provide indications of how well on the average a school is doing in those categories in which its performance is assessed against other similar schools. It is, however, obvious that two systems could have equal predicted and actual output means yet be quite different in their effect on superior students and/or below average students.[7] For example, a gain in mean score at school A could be due to large increases for below-average students with relatively smaller increases for above-average students. An equal mean gain at school B could result from large gains for above-average students and small gains for below-average students. One way of learning about such differences is to set aside the comparison based on means and examine how above- and below-average students perform in these two schools.

For example, the above two schools with equal performance indices based on means would have quite different performance indices if they were derived from within-school-regression-equations of individual student outputs on corresponding individual student inputs. School A, with relatively larger gains for below-average than for above-average students, would have a regression line with a relatively flat slope whereas school B, with relatively smaller gains for below-average students than for above-average, would have a regression line with a steep slope. Figure 4 provides an illustration of how the within-school regression lines might appear for these two schools. Schools A and B both have a mean input score of 65 and a mean output score of 75 (i.e., the point at which the two lines in Figure 4 intersect).

---

[7] Goodman, op. cit., found this to be the case with the first QMP data. See pages 19 and 20 of his 1959 report.

However, it will be noted that students from both schools who have an input score of 40 (below average) have a predicted output score of 70 at school A and 55 at school B. Input scores of 40 and the corresponding predicted scores of 70 at school A and 55 at school B are indicated in Figure 4 by letters a and b respectively. At the other extreme, all students with an input score of 90 (above average) have a predicted output score of 80 at school A and 95 at school B. This fact is indicated in Figure 4 by α and β. Note that although students at both schools show the same mean, gains (from an input of 65 to an output of 75), students with an input score above the mean of 65 score higher at school B than do those above the mean at school A, whereas the converse is true for students with input scores below 65.

The regression coefficient is equal to the slope of the regression line, and thus it provides an index of any difference in impact a school may have on its above-average as compared with its below-average students. The regression coefficients for the example in Figure 4 are .2 for school A and .8 for school B. These regression coefficients, coupled with the performance indices based on school means provide additional information for distinguishing between these two schools.

The actual estimation of the regression coefficients would involve a large data processing job since it requires the use of individual student records and computations within each school system. However, it could provide the information necessary for discovering important differences between otherwise similar schools.

In order to facilitate the interpretation of the regression coefficients, they would be converted into "slope indices." For each output measure the regression coefficient for a school would be compared to the corresponding

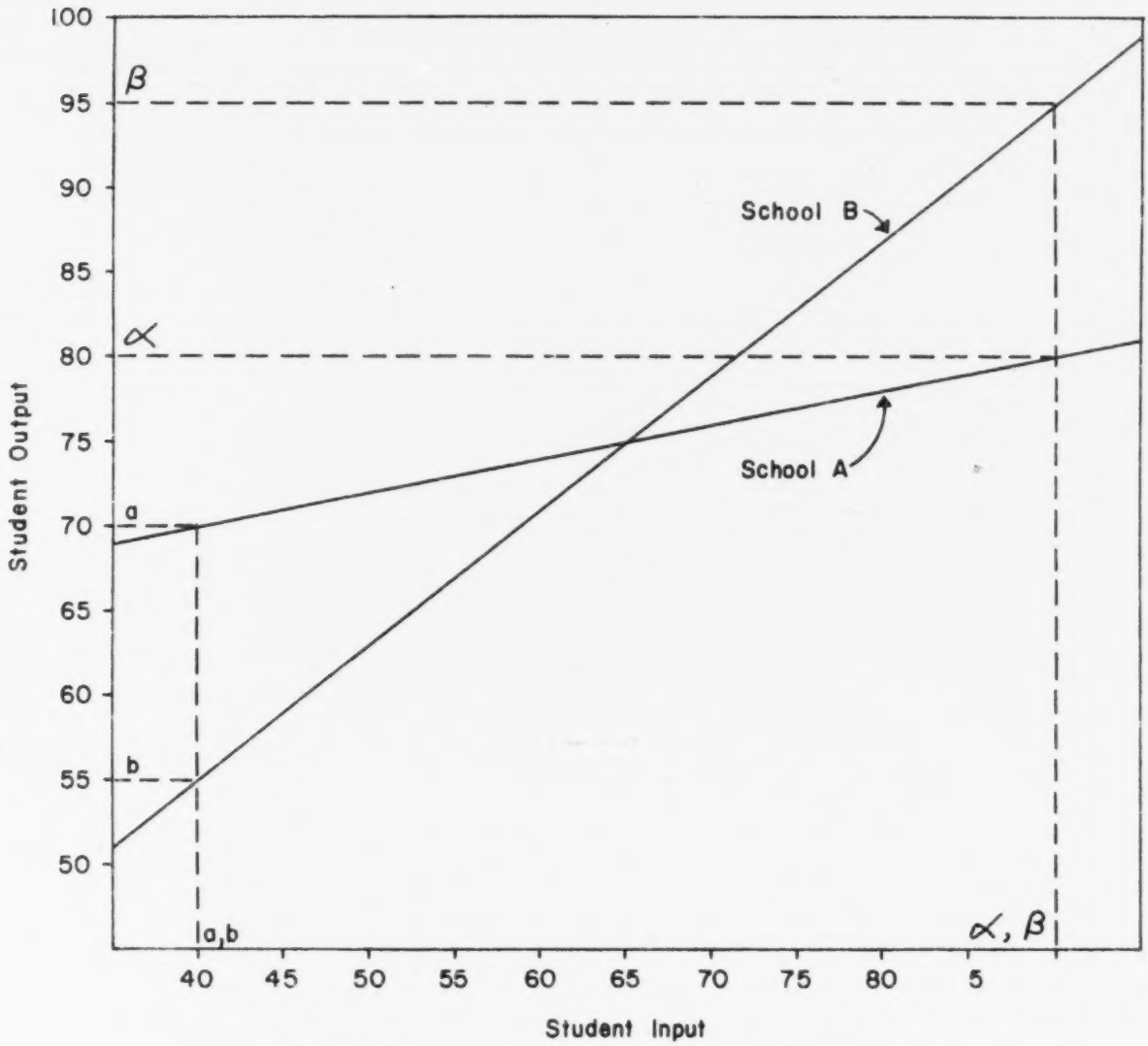

Figure 4

Illustration of Within School Regression Lines for Two Schools with equal Performance Indices Based on Means

coefficients for other schools. Schools with regression coefficients among the 20 per cent that are the smallest would receive slope indices designated C, schools with coefficients among the 20 per cent that are largest would receive slope indices designated A, and all other schools would receive slope indices designated B.

Alternative Procedures

The approach presented in the preceding pages is certainly not the only reasonable one. It is our judgment, however, that the procedures outlined above have the greatest likelihood of proving fruitful. A major aspect of the recommended pilot study would be the investigation of a number of alternative procedures.

An alternative to the longitudinal procedure which is planned could entail a cross-sectional analysis which uses surrounding condition variables alone to make adjustments in output. A cross-sectional approach might also use measures of current third-graders to make adjustments on the measures of current fifth-graders and so forth. A cross-sectional approach is not considered to be desirable in and of itself, but rather it would be desirable only to the extent that it provided adequate approximations to the theoretically preferable longitudinal approach. The obvious advantage of a cross-sectional approach is that it is operationally much simpler and less costly.

There are also alternatives to the proposed basic regression analyses. It might be easier to interpret some sort of difference score which would be a more obvious measure of change than the residual scores that are now being used to determine performance indices. Alternatives of this kind would be considered in the pilot study.

Another example of alternative procedures arises in the consideration of the way in which possible differential effectiveness of a school for various subgroups

of students might be discovered. Instead of the within-school regression analyses which are suggested, it would be possible to approach this problem in essentially the same way as the analysis of school means. By substituting various points in the score distribution of a school, like the 20th and 80th percentiles, for the school means, performance indices could be developed for the effectiveness of a school with the below-average and the above-average students separately, and in addition to the performance indices computed for the average. Once again, the pilot study would provide an opportunity to compare the different approaches.

V. UTILITY OF THE INDICATORS

Their function

What the performance indicators give us is a series of rough estimates of how an educational system is doing in comparison with other systems working under presumably equivalent advantages or handicaps. The fact that there will be a matrix of indicators will help to show at what points in its educational program a system may be strong or weak in the opinion of the administrators and policy-makers responsible for it. That is, they will suggest where new effort and money might produce pay-off in improving the development of the youngsters in the system. The indicators themselves, however, will not suggest specific remedies or the cost of the remedies, whether cost be calculated in terms of money or of opportunities foregone.

The remedies are to be found in two sets of independent variables that in the derivation of the performance indicators were purposely omitted: measures of educational processes and measures of the more modifiable conditions in which a school system operates. The function of these measures in the assessment of the schools would be precisely to provide clues to the steps that might be taken to improve their performance. By examining the specific differences in modes of operation between the top and bottom school systems in any category, it should be possible to get some good approximations of what ought to be done to move the bottom schools closer to the top ones.

The indicators in Figure 3 illustrate how 91 school systems are performing in respect to reading output at grade 6 relative to both their own predicted output and the performance by other systems.[8]

---

[8] In practice, there would be many more than 91 school systems.

Suppose that the following differenses are found among Systems A, B, and C and System D.

(a) Systems A, B, and C all have a program of prescribed summer reading, while System D does not.

(b) A, B, and C work intensely with their local public library to encourage voluntary reading through extensive distribution of paperbacks, week-end book conferences, and the like; D has no contact with the public library system.

(c) Systems A and B have instituted a tutorial program for non-readers in which senior honor students are employed as tutors. Neither of the other two systems has such a program.

(d) System A has organized a parent-teacher study group to explore ways of encouraging more and better reading in the home. There is no such program in B, C, or D.

There is, of course, no guarantee that instituting these programs in System D would improve its reading output to the point where the performance index would move up from 1 to 5. Nevertheless, the strong presumption would be that by adopting the modes of operation which are being successfully used by systems more or less like itself, it would at least begin to show some improvement.

The great utility of educational performance indicators coupled with good measures of educational process and the modifiable conditions of learning is that they highlight the specific steps that specific schools and school systems might conceivably take to help their pupils grow and find themselves. They should also have the effect of reducing to some extent the guesswork in allocating resources and deploying educational personnel so as to maximize the effectiveness of the system.

Hypothetical Example

The basic results that would be provided for a given school would consist of two matrices of indices. The first of these would be a matrix which reported the

performance indices for each phase of education in each of the major categories of performance. The second matrix would be exactly the same in form but would report the slope indices in each category for each phase.

A matrix of performance indices and a matrix of slope indices are presented in Figure 5 for an hypothetical school. The matrices are essentially the same as the illustrative matrix presented in Figure 2 except the number of phases has been reduced to simplify the presentation.

The most obvious feature of the hypothetical performance indices in Figure 5 is that the school is performing exceptionally well in the cognitive area at all three phases of education under consideration. It is also apparent that it is performing considerably less well in the other performance categories, and especially so in the higher grades. Such a pattern of performance indices would strongly suggest that its excellent performance in the cognitive category is being purchased at the expense of its performance in the other categories. Whether this is a desirable state-of-affairs for the school in question would depend upon the specific set of goals of that school as defined by the community in which the school is located. The important point is that the performance indices provide an objective means of determining how effectively the particular goals of that school are being achieved.

The slope indices which are provided in the lower part of Figure 5 indicate that the differential effect of the school for above- and below-average students is about typical of all other schools except for the physical and cognitive categories. In the physical category, there is consistently a slope index of C which indicates that the school tends to be relatively more effective with below-average than with above-average students to a greater extent than is typical of other schools. The opposite situation obtains at the two higher phases of education in the cognitive category, i.e., the slope index is A. Thus at phases 5 to 7

Figure 5

Performance and Slope Indices for

an Hypothetical School

Performance Indices

Phases of Education by Years in School	Major Categories					
	A Physical	B Cognitive	C Social	D Moral	E Vocational	F Recreational
Year 9 to Year 11	3	5	2	2	1	1
Year 5 to Year 7	3	5	3	3	1	2
Year 1 to Year 3	3	5	3	3	3	3

Slope Indices

Year 9 to Year 11	C	A	B	B	B	B
Year 5 to Year 7	C	A	B	B	B	B
Year 1 to Year 3	C	B	B	B	B	B

and 9 to 11 this school is relatively more effective with its above-average than its below-average students, as compared to other schools in its set.

The results presented in Figure 5 can be further simplified by graphing the performance indices and simply noting the corresponding slope indices. The same information is presented in Figure 6A as was presented in Figure 5, but the results are presented here in graphic form. In Figure 6A it can be seen that there are three bars for each performance category: one for each of the three phases of education. The performance index is indicated by the length of the bar and the slope index is indicated by cross-hatch marks for a slope index of A, diagonal lines for a slope index of B, and no lines for a slope index of C.

Figure 6B presents analogous data for a second hypothetical school (school Y) from the same set of schools. As was the case for school X, school Y has a performance index of 3 for all three phases of education in the physical performance category. By way of contrast, however, the slope indices in the physical category of school Y are all A, whereas, they are all C for school X. This is a situation that is similar to the one portrayed in Figure 4. The student who has an above-average input score in the physical category probably would gain more at school Y than at school X, while the below-average student probably would gain more at school X than at school Y. Which of these two situations is preferable depends upon judgments that should be made by citizens of the community.

School Y has performance indices which are considerably lower than those of school X in the cognitive and moral categories, but the performance indices of school Y are higher than those of school X in the social and recreational categories. Once again, the pattern to be preferred depends upon value judgments that should be made by the citizens of the community. Therefore, data such as these could, one would hope, induce school-community dialogues about those goals of education that are desired by the given community.

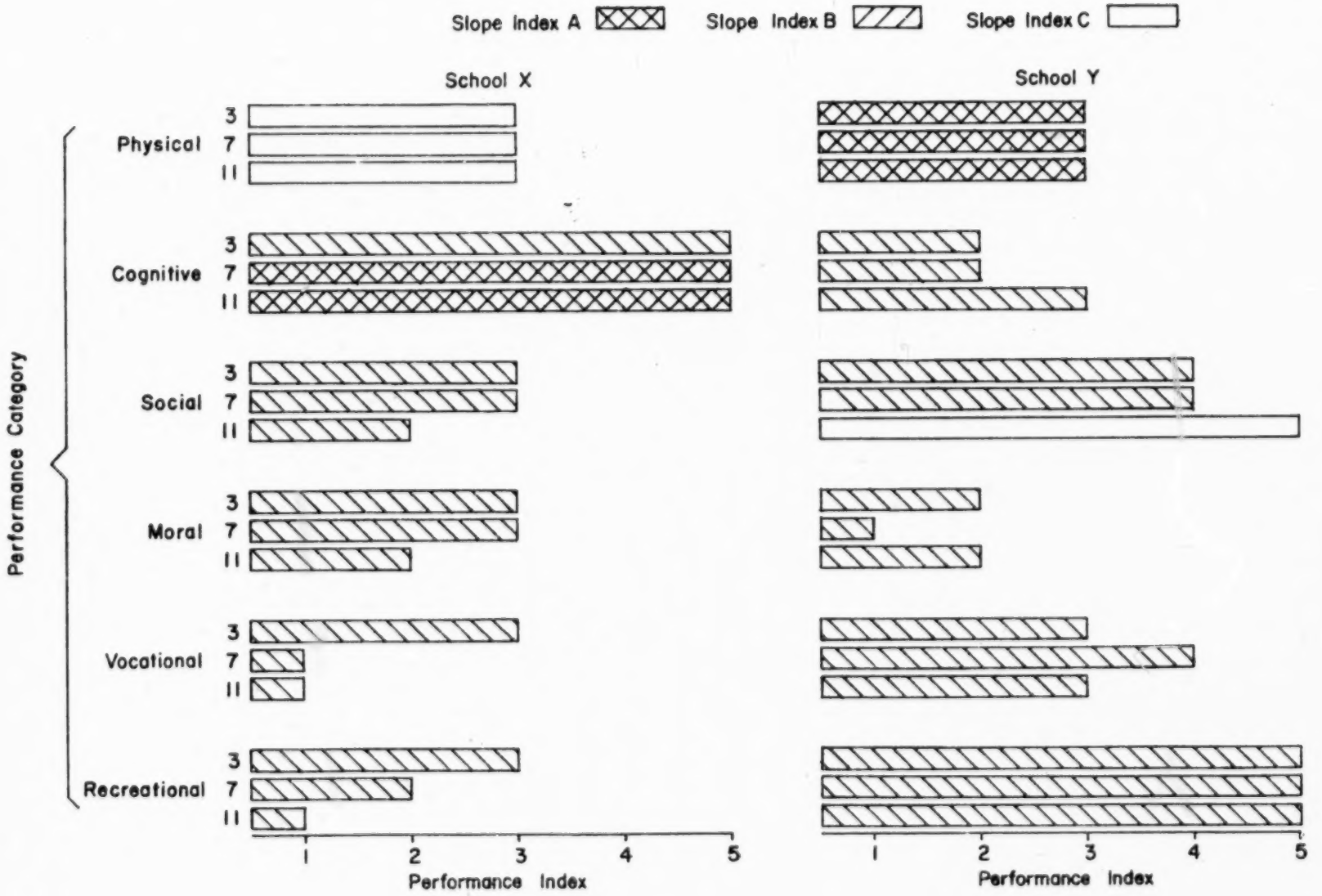

Figure 6A
Profiles of Performance Indices
at Hypothetical School X

Figure 6B
Profiles of Performance Indices
at Hypothetical School Y

VI. QUESTIONS TO BE INVESTIGATED BY THE PILOT STUDY

The proposed procedure for developing educational performance indicators rests upon a number of underlying assumptions whose validity can be properly evaluated only in the light of empirical results. The pilot study should provide information crucial to the evaluation of these underlying assumptions. Its results should also form the basis for modifying the procedure to be adopted in developing educational performance indicators for actual use by the schools.

The study would be addressed to the following questions:

1. Is the linear regression model, described in Appendix B, appropriate for predicting output means from input means and measures of surrounding conditions? There has been insufficient work with this type of data to make clear the nature of the relationships to be expected.

2. Is there enough residual (i.e., unaccounted for) variance among school systems with equal predicted outputs to make worthwhile the attempt to attribute any significance to the actual differences? If the variance of the actual outputs around their predicted outputs (i.e., the residual variance) is approximately equal to the variance that might be expected from measurement errors alone, then there would be little point in attempting to attribute any significance to the differences between predicted and actual outputs. On the other hand, if the residual variance is considerably larger than the amount of variation that could be explained on the basis of measurement errors, it is reasonable to search for systematic differences between school systems having outputs that are grossly overpredicted and those having outputs that are grossly underpredicted.

3. What is the best procedure for developing indices that will reflect the effect of the school system on the top and bottom students in the system?

Two systems could have equal predicted and actual output means yet be quite different in their effect on superior students and/or below-average students. For example, a gain in mean score at one school could be due to large increases for below-average students with relatively smaller increases for above-average students. An equal mean gain at a second school could result from large gains for above-average students and small gains for below-average students. How can we determine what has contributed to these gains in school means? The primary approach to this problem would be to develop an index that would illuminate the differential performance of above- and below-average students within a school. A slope index derived from within-school regression analyses would, one hopes, provide the needed information. This approach and others need to be evaluated.

4. How stable are the performance indices from one sample of students to another within the same school at the same point in time? This question is closely related to the second question above, but is more specifically aimed at the actual determination of the width of the bands to be used in converting deviations from the regression line to performance indices. A discussion of the way in which the bands are derived can be found in Appendix B.

5. Is it necessary to require that the performance indices be derived from repeated testing of the same students? The current formulation requires that the input and output measures used in developing the performance indices be based on the same students. From an operational point of view, this is a demanding requirement; if essentially the same results could be obtained by a less stringent requirement, considerable savings could be achieved. Two alternative procedures would be investigated. The first procedure would retain the requirements of longitudinal data but would not be limited to those students for whom both input and output measures are available. Under this plan input statistics would be based on all students for whom input measures were available. Similarly output statistics would be based on all students for whom output

measures were available. The second alternative procedure simply involves the use of cross-sectional data in place of longitudinal data. For example, if students in grades 1 and 3 were tested at the same point in time, the grade 1 test results would be used as the input measures from which the output measures (i.e., grade 3 test results) would be predicted. In this case the two groups would be treated, for predictive purposes, as if they were one.

6. Would the prediction that can be achieved from input variables alone be substantially improved by including measures of hard-to-change surrounding conditions?

7. Would the prediction that can be achieved from measures of hard-to-change surrounding conditions be substantially improved by including input variables? This question relates to question 5 because if measures of student input do not add substantially to the prediction, then it may be quite reasonable to ignore the measures of student input -- a situation which would eliminate the need for longitudinal data.

8. For sets of school systems with similar predicted outputs, can systematic differences be identified between systems with high performance indices and systems with low performance indices? This question is directed at one of the fundamental justifications for the development of performance indices, namely, the identification of specific steps that might be taken within a given school which perhaps may lead to improvement.

9. Would knowledgeable educators (e.g., State supervisory personnel, participants in the Cooperative Review Service) make judgments about the degree of similarity or differences among schools in the opportunities they provide to students such that their judgments would correspond to those obtained by the series of predicted outputs?

In addition to answering the foregoing questions, the pilot study should result in valuable experience in coping with some of the practical problems that must be faced before placing a system for the development of performance indicators on an operational basis. One major problem would be to develop interpretive material that would make the meaning of the indices readily understandable by school personnel. One approach to this problem might be to develop a standard series of verbal statements from among which a computer could select those appropriate to each school profile.

VII. RECOMMENDATIONS

The purposes of the pilot study are to try out, evaluate, and modify the proposed procedures for developing educational performance indicators. Toward these ends, answers will be sought to the questions which were raised in the preceding section of this report.

The General Plan

The pilot study is divided into two phases. The first phase would be developed around the longitudinal student characteristics data available through the Quality Measurement Project (QMP). The second phase would depend upon the Pupil Evaluation Program (PEP) for data on student characteristics and would also rely heavily on data available from the Basic Educational Data System (BEDS) for measurement of the surrounding conditions and the educational process variables.

Each of the two primary data sources (the QMP and the PEP) has important advantages over the other source that consequently makes it desirable to use both in the pilot study. The QMP data include measures of more student characteristics and are available for more grade level groups. On the other hand, the PEP data are current and thus more relevant to present educational problems. Another prime advantage of the PEP data is that they can be associated with the rich information on surrounding conditions and educational processes that became available for the first time through the Basic Educational Data System in the fall of 1967.

Thus, the QMP data are viewed as providing a better source for investigating important theoretical and technical questions concerning the basic procedure. More specifically, the QMP data would permit more comprehensive answers to

questions 1 through 6 than would be possible with the PEP data. Given that the indices are theoretically and technically sound, however, the PEP data coupled with data obtained from BEDS are far superior for evaluating the potential usefulness of educational performance indices. The PEP would also be logical framework for the next steps in the development of performance indices following the pilot study.

Phase One--Study Involving QMP data

Summarization of previous QMP results. It is quite likely that previous research with the QMP data would make some of the analyses described below unnecessary and it would be useful to re-examine these data for this reason. Certainly much of the research with QMP data has a direct bearing on one or more of the above nine major questions to be investigated, and in some cases the previous analyses may provide a sufficient answer making further analyses for that question unnecessary. For example, in the summary of QMP data and research provided by Dr. Robertson on August 1, 1967, correlational and regression analyses using 1965-66 school means are listed. These results might be quite relevant for questions 1, 6, and 7.

In view of the fact that the QMP originated ten years ago and has generated many analyses and reports both of a formal and an informal nature, we think that it would be of great value to combine in the form of a single report those results that are thought to be most relevant for the above questions. Such a document undoubtedly could provide partial answers to some of the questions and might provide completely sufficient answers in some instances. In any event, a report summarizing relevant results would make it possible to determine precisely those analyses that still may be needed, thus making it possible to update the specification of procedures.

The analyses of the QMP data that are described below, or at least some of them, may be unnecessary. They are presented here on the assumption that at least some of them will be necessary and even if this should prove to be incorrect, their specification will make more feasible a summarization of the relevant results of previous studies as they pertain to the questions to be investigated.

The data for new analyses--if needed. This phase of the study would involve three grade level groups, namely, QMP students who were in the 4th, 7th and 10th grades in 1957-58, in 90 odd school districts. The student characteristics data obtained in 1957 would be used as the measures of student input and the corresponding data obtained in 1959 would be used as the measures of student output. The measures of surrounding conditions would be limited to existing data of record and the student reports of father's occupation.

A more detailed list of the variables in the three major categories is given below. Groups A, B, and C refer to the groups of students that were in grades 4, 7, and 10 respectively in 1957. The grade levels at time of input and output measures are shown in the table.

Group	Input Measures 1957	Output Measures 1959
A	4th Grade	6th Grade
B	7th Grade	9th Grade
C	10th Grade	12th Grade

Measures of Output (1959 Data) for Groups A and B

(Iowa Tests of Basic Skills)

1. Vocabulary
2. Reading
3. Language
4. Work Skills
5. Reading
6. Composite

Measures of Output (1959) for Group C

(Iowa Tests of Educational Development)

1. Basic Social Concepts
2. Background in Natural Sciences
3. Correctness and Appropriateness of Expression
4. Ability to Do Quantitative Thinking
5. Interpreting Reading Materials in Social Studies
6. Interpreting Reading Materials in Natural Sciences
7. Ability to Interpret Literary Materials
8. General Vocabulary
9. Uses of Sources of Information
10. Composite

Measures of Input (1957 Data for Groups A and B

(Iowa Tests of Basic Skills)

1. to 5. Same as output measures
6. General ability measure - (Lorge-Thorndike Intelligence Test)

Measures of Input for Group C

(Iowa Tests of Educational Development)

1. to 9. Same as output measures
10. General ability measure - (Lorge-Thorndike Intelligence Test)

Measures of Surrounding Conditions for All Three Grades

1. Community Type
2. Father's Occupation (coded as high, middle or low SES)
3. Property Valuation Behind Each Pupil
4. School Tax Rate
5. Number of Professional Personnel per Pupil
6. Number of Publicly Owned Instruction Rooms
7. Attendance
8. Median Degree Status of Teachers
9. Median Years Experience of Teachers
10. Median Teacher Salary
11. Proportion of Teachers with Tenure
12. Enrollment

The procedure for new analyses if needed. The individual student data from 1957 and from 1959 are now on a single punch-card for each student. The first task would be to put on magnetic tapes all the individual student data and the measures of surrounding conditions for each group in each school. This work should begin in March 1968, and by the end of April the data tapes should contain three points in the distribution for each school: the 20th and 80th percentile, and the mean. For the purpose of cross-validation later on, these three points in each of the distributions would be determined on random halves of the students in each grade in each school. Linear regression analyses of each output measure on its corresponding input measure would be performed within each school and the regression coefficients put on the data tapes. Once the data had been thus prepared, the following analyses would be performed:

Question 1 (How appropriate is the linear regression model?) would be approached by producing and inspecting scatter plots of each of the outputs with each of the predictor variables. If large departures from linearity are observed, either a method of transforming variables or non-linear models would be used.

For each of the 22 output measures (6 for group A, 6 for group B, and 10 for group C) the mean output score would be regressed on the input means and measures 1, 2, and 3, of the surrounding conditions.[9] Several stepwise regression analyses (and possibly other methods for reducing the number of predictors) would be performed for each output. One set of analyses would be based only on students for whom measures are available at both points in time (longitudinal data -

[9] These are considered the "hard-to-change" conditions. The influence of the remaining nine "modifiable" conditions would be investigated subsequently to determine the extent to which they might reduce the remaining variance within arrays. No preliminary factor analyses of these measures would be performed since they are already relatively few in number.

same students only). The same analyses would also be performed based on the means of all students who have scores on the variable in question (longitudinal data - all students). A comparison of the results of these analyses would provide answer to question 5 (Is it necessary to base analyses on repeated measures of the same students?).

Stepwise regressions would also be performed with all variables free to enter and with measures of the hard-to-change surrounding conditions allowed to enter only after all input variables had entered the equation. This set of analyses would permit tentative answers to question 6 (Do measures of surrounding conditions improve the prediction possible from inputs alone?).

Conversely, the regression analyses would be computed with input variables allowed to enter only after all hard-to-change measures of surrounding conditions had entered the equation. These analyses would be used to answer question 7 (Does the inclusion of measures of student input substantially improve the prediction of output that can be achieved from measures of surrounding conditions alone?).

The meaningfulness of the departures of actual means from predicted means would be investigated in three ways. First, the variance of the simple differences between school output means and the corresponding input means would be computed and the hypothesis that this variation is due only to errors of measurement would be tested by means of an F-test. Similarly, an F-test would be used to test the hypothesis that the standard error of estimate is only measurement error. These analyses are directed at question 2 (Is the residual variance of sufficient size to give meaning to the differences between actual and predicted outputs?). Following these analyses the performance indices would be computed for each school system, using each half-sample in turn. The regression weights developed in the

first sample would then be applied to the hold-out sample within each system and the performance indices computed for the hold-out samples. The indices for the two samples for each school would then be compared in order to get an indication of the stability of the indices, and thus provide a tentative answer to question 4 (How stable are the indices?).

Two additional regression analyses for each output would also be performed in similar manner using the 20th and 80th percentiles respectively as the input and output measures of a school system. For these analyses only matched cases would be used. These analyses are directed at question 3 (How should indices be developed to reflect the performance of the system in regard to above-average students and below-average students?). The results of these analyses would be compared to the results based on the within-school regression coefficients.

Surrounding condition variables 4 through 12, plus possible additional measures of teacher variables and any measures of educational process that could be obtained for QMP school systems, would be used to compare systems that have low performance indices with systems that have approximately equal predicted outputs but high performance indices. This step is in response to question 8 (For sets of school systems with similar predicted outputs; can systematic differences be identified between systems with high and low performance indices?). However, a better and more complete answer to this question would be obtained in Phase II using PEP and BEDS data.

Phase Two -- Study Involving PEP and BEDS Data

The Data. The only group of students for whom current longitudinal test data are available is the group of students that was first tested in the first grade in 1965 and then tested again as third graders in the fall of 1967. The prime source for measures of surrounding conditions and educational process would be the

Basic Educational Data System (BEDS). The measures that would be used in this phase of the study are listed below under the four major variable categories.

Output Measures (1967 - grade 3)

1. Reading Achievement
2. Arithmetic Achievement

Input Measures (1965 - grade 1)

1. Readiness Test
2. General Ability

Measures of Surrounding Conditions[10]

1. Population density (CU)
2. Property valuation (CU)
3. School tax rate (SM)
4. Number of instructional personnel (SM)
5. Proportion of professional classroom personnel who are certified (SM)
6. Proportion of professional non-classroom personnel who are certified (SM)
7. Ratio of professional personnel to enrollment (SM)
8. Median degree status of classroom teachers (SM)
9. Median years experience of classroom teachers (SU)
10. Median salary of classroom teachers (SM)
11. Median years experience in present assignment for classroom teachers (SU)
12. through 15. Same as 8 through 11, but for professional non-classroom personnel (SM)
16. Median proportion of day administrators devote to teaching (SM)
17. Median degree status of administrators (SU)
18. Median years experience administrators have in education (SU)
19. Median years experience administrators have in administration (SU)
20. Median years experience in present position for administrators (SU)
21. Proportion of professional personnel that are non-white (SM)
22. Proportion of students that are non-white (SM)
23. Attendance (SU) (by district only; by fall 1968; by school)
24. Number of classrooms (SM)
25. Enrollment (SM)

---

[10]The letters in parentheses have the following meanings: C=community condition; S=school condition; U=unmodifiable condition; M=modifiable condition. This coding will require checking by others for its reasonableness.

Measures of Educational Process

1. Participation in regional programs
2. Programmed learning
3. Computer-assisted instruction
4. Other types of independent study
5. Closed circuit television
6. Open circuit television
7. Ungraded continuous progress (elementary level)
8. Curricular innovations
9. Flexible or modular scheduling
10. Pre-kindergarten program
11. In-service teacher education
12. Program for implementing intergration and intergroup relations
13. Specially funded ESEA Title I project
14. Specially funded ESEA Title II project
15. Specially funded ESEA Title III project
16. Specially funded foundation project
17. Specially funded project: other sources
18. Employment of consultants: administration
19. Employment of consultants: facilities
20. Employment of consultants: in-service teacher education
21. Employment of consultants: public relations
22. Attendance service
23. Guidance
24. Health service
25. Psychological service
26. Social work service

Measures 22 through 26 will have to be obtained from professional personnel forms)

Procedure. The sample for this phase of the study would consist of 200 schools so chosen as to provide three widely different groups -- each group as homogeneous as possible with respect to such hard-to-change conditions as urbanness, ethnicity and wealth of community, average socio-economic status of pupils, and mobility of student body. Within each group we would hope to get as much variation as possible in such modifiable variables as educational effort of the community (ratio of school tax to taxable wealth), educational processes, and the like.

Since the PEP data are not gathered centrally on an individual student basis, the input and output data would have to be obtained from local school systems. It is proposed that local schools selected for the study arrange to have individual

scores from the 1965 readiness test, general ability scores, and the 1967 test measures recorded on roster sheets and returned to the project staff. It is assumed that the roster sheets could be designed so that the recorded information could be read by an optical scanning device.

Once the individual student data had been put on magnetic tape, the next major operation would be merging the measures of surrounding conditions and educational process variables with the PEP data.

When the data had been merged, the performance indices would be developed for each of the output measures. The predictors to be used in deriving the performance indices would be the readiness test and measures of the "hard-to-change" surrounding conditions. The decision that a given surrounding condition is "hard" to modify is, at least to some extent, a policy decision and should be made by policy makers in the State Education Department. The actual form of the regressions, the summary statistics (e.g., means, upper and lower percentiles, within-school regression coefficients), and the determination of the ranges of the deviations from the regression line of the several levels of performance indices (i.e., the width of the bands) would be determined in light of the results of Phase I. Systematic comparisons would be made among systems with similar predicted output that have obtained either "high" or "low" indices. Measures of surrounding conditions and of educational process that were not included in the prediction equation would be identified and used in making these comparisons.

Attempts to reduce the number of variables to be used for making comparisons would be made via cluster analysis and/or factor analysis. As one means of comparison, regressions of residual output scores on these variables would be computed for groups of schools with approximately equal predicted outputs. Systematic differences that occur among systems could then be investigated as events

which might underlie the differences in system performance. These analyses are intended to provide answers to question 8 (For sets of school systems with similar predicted outputs, can systematic differences be identified between systems with high and lower performance indices?).

Question 9 (Do judgments of the similarity of school systems by knowledgeable educators agree with the similarity of predicted outputs?) would be approached by having supervisory personnel make judgments of the similarity of groups of school systems, and then these judgments would be correlated with the predicted outputs.

Implementation

In the interest of implementing the pilot study we recommend that the Department Committee for the EPI Project meet as soon as possible to:

1. Review progress to date.
2. Determine what analyses of QMP data have been completed that provide answers to the basic questions which are listed above.
3. Specify the operational and management procedures to be employed in performing the additional analyses of QMP data.
4. Specify analogous procedures for the analyses of the PEP and BEDS data.

VIII. LONG RANGE PLANS

Developing an Operational System

If the proposed pilot study for 1967-68 were to provide enough favorable answers to suggest that an operating program is practicable, we see the following possibilities for the future:

In 1968-69, a limited operational program based on analysis of PEP longitudinal data that would presumably become available in the fall of that year at grades 6 and 9.

In 1969-70, upward extension of the PEP to grades 11 and 12 so as to provide some kind of tie-in with the Regents Examination Program and the Scholarship Examination Program for the purpose of getting more adequate measures of system performance at the secondary level.

In 1969-70, introduction of noncognitive measures of pupil performance at all levels in order to begin to provide some indication of the impact of the schools and the community on the personal and social development of pupils.

In 1970-71, incorporation of the proposed regional data processing centers into the EPI system so as to facilitate data collection, analysis, and reporting in greater detail and depth.

Continuing Modification and Simplification

The development of an operational system of performance indicators would require continuing modification and simplication. A number of questions about the procedural details of the indices could be answered, at least in part, by the pilot study. Answers to these questions would form a basis for the modification

of the procedure. As the system develops, however, there would be a continuing need for investigations to answer new questions and provide a means of monitoring the workings of the system.

As the system develops there also would be an increasing need to simplify the presentation of information provided by the system so that it could be readily understood and used by the practitioner. This does not imply that the derivation of the indices would necessarily become simple. On the contrary it might become more complex, but no matter how involved the actual derivations might become the end results must be presented in a form that would be readily interpretable while still providing the most accurate representation possible.

Prospect for Cost-Benefit Analyses.

One of the possible values of educational performance indicators, as we conceive them, is that they might eventually provide a basis for some genuine cost-benefit analyses. If our notions prove out in the proposed pilot study, we see the indicators as constituting the educational "benefit" side of the cost-benefit equation.

We wish to emphasize, however, that the whole conception of cost-benefit analysis as applied to education is still in its infancy. There are tough theoretical problems in the analysis of social systems as compared, say, to missile systems, that are still not solved. The whole field of program budgeting in education is still not well understood and rarely attempted in any serious fashion. We do not wish to raise any premature hopes therefore that the investigations proposed for developing educational performance indicators would result a year hence -- or even five years hence -- in a rigorous decision system for the allocation of educational resources in New York State or in particular school systems within the State.

On the other hand, we are equally convinced at this point that an empirical study to develop educational performance indicators of the type we are suggesting is an absolutely necessary first step toward a practical program-planning budgeting system.

Experimental Inquiry

If educational performance indicators are to become a useful application of technology to the problems of education, then the thoughtfulness with which this technology is applied ought to be enhanced by a corresponding increase in our knowledge of why the indicators subsequently appear as they do.

The Indicators represent the tying together of student performance and other measures through regression analysis so that judgments can be made about the effectiveness of schools. Thus, the indicators are a kind of systematically descriptive first approximation of what factors are likely to be associated with student growth. This is valuable information in and of itself if we do not go beyond it to make unfounded cause-and-effect statements about how and why student performance takes the forms it does. Only when we can show that certain changes in antecedent conditions (student input, school and community factors, etc.) are associated with changes in the value of a given performance measure will we be able to infer cause and effect with more assurance.

What is needed, then, is not only the results of the regression analyses expressed as performance indicators but also experimental manipulation of treatments and conditions to see what factors are most likely to be affecting pupil development. It is here where performance indicators derived from the analysis can be of great value in helping specify the treatments and conditions to be experimentally investigated. They can suggest those factors which are likely to prove amenable to empirical description and treatment as independent events that can then be manipulated

in order to observe change in student performance. It is through the experimental manipulation of carefully specified antecedent variables that we can hope to gain more than unverified hunches about why and how students perform as they do and what types of school programs may be most effective in helping particular types of students to learn.

For example, after performance indicators point to differences among schools in the output performance of students, clearly specified "process" and "surrounding condition" variables can be treated as independent or antecedent events from which to suggest and test hypotheses about their likely effect on student performance, treated as the dependent event. The data provided by the performance indicators would underscore the need to define formally and test empirically what personal, school, and community characteristics are antecedent to the observed outcome differences among schools described by the indicators. The indicators would provide a rich store of information from which answerable questions about the possible sources of influence on student performance may be raised.

Appendix A

Factor Analyses

Maximum likelihood procedures using Joreskog's[1] computational technique would be used for all of the factor analyses. Joreskog's program permits the extraction of any number of factors, and each factor matrix is in turn rotated via Kaiser's varimax method.[2] Hypotheses about the number of factors are tested by means of a chi-square test based on the likelihood ratio technique. These tests are based on the assumption that the variables have a multivariate normal distribution.

The chi-square tests are addressed to the following series of questions: Is the correlation matrix significantly different from the identity matrix? If so, is there a factor, f_1, such that the partial correlations between pairs of variates are not significantly different from zero after the effect of f_1 has been removed? If not, are there two factors, f_1 and f_2, such that the partial correlations between pairs of variates are not significantly different from zero after the effects of f_1 and f_2 have been removed, and so on?[3] Since the sample of students will be quite large, a significance level of .01 will be used for the acceptance of a factor.

In the analysis of the output variables, the number of factors, K, that are significant at the .01 level would determine the number of measures

---

[1] K. G. Joreskog. UMLFA--A computer program for unrestricted maximum likelihood factor analysis. Research Memorandum 66-20. Princeton, N. J.: Educational Testing Service. Revised Edition. 1967.

[2] H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

[3] D. N. Lawley and A. E. Maxwell. Factor Analysis as a Statistical Method. London: Butterworth, 1963.

to be retained. The varimax rotation of the K significant factors would be used to decide upon the K measures to be retained. The output measure that has the highest loading on the first factor would be taken as the first measure and eliminated from further consideration; then the output measure that has the highest loading on the second factor would be taken as the second measure, and so on. The factors would be ordered according to the proportion of variance accounted for (i.e., the sum of the squared factor loadings) before measures were selected.

There would be two exceptions to this otherwise mechanical procedure for reducing the number of variables. First, for an output measure to be retained when considering a given factor it must have a loading of at least .40. If the highest loading for the unselected variables on a factor were less than .40 then no variable would be selected for that factor. If this should happen it might indicate that a new measure needs to be constructed which would be more highly related to the unrepresented factor. The second exception is that a variable would be retained as an output measure if it had a uniqueness of .40 or larger whether or not it would be accepted on the basis of the above criteria. The choice of .40 as the cutoff point is admittedly somewhat arbitrary. It is our judgment, however, that a factor loading less than this is insufficient to justify using the variable as a substitute for the factor, and that a uniqueness larger than .40 indicates potentially important variance that is not common to the other variables.

In the case of the factor analyses of the hard-to-change surrounding condition variables, the modifiable condition variables, and the educational process variables, the factor analysis procedures would be the same but the method of reducing the number of variables would be different. As above, the significant factors would be rotated by a varimax rotation and ordered according

to the proportion of variance accounted for. An attempt would then be made to interpret each of the factors. If interpretable the scores on the measures contributing to the interpretation of a factor would be summed to form a single score. There would be a restriction that a measure be part of only one score. Once again, a measure would be retained if it had a uniqueness of .40 or larger.

Appendix B

Regression Analyses and Conversion of Deviations to Performance Indices

For each regression analysis, the dependent variable would be the school mean of the output measure being considered. The independent variables would be of two kinds: means of the measures of student input characteristics and measures of hard-to-change surrounding conditions. Let O_{cgs} be the mean output in category c, at grade level g at school s; $I_{cg's}$ be the mean input in category c, at grade level g' and school s; and $S_{j(g'-g)s}$ be the mean hard-to-change surrounding condition j obtaining between years $g' - g$ at school s.

The predicted output, designated $\hat{O}_{cgs}$, in category c, at grade level g, in school s is then given by:

$$\hat{O}_{cgs} = \sum_{c=1}^{m} b_{I_c} I_{cg's} + \sum_{j=1}^{p} b_{s_j} S_{j(g'-g)s} + a$$

where the b's are the regression coefficients, a is a constant, m is the number of performance categories, and p is the number of hard-to-change surrounding condition measures.

The difference between the actual output mean O_{cgs} and the predicted output mean $\hat{O}_{cgs}$ would be used to determine the actual performance indices. First, the standard error of a mean, SEM, would be computed as follows:

$$SEM = \overline{SD}/\sqrt{\bar{n}}$$

where $\overline{SD}$ is the average of all the within school standard deviations for the output measure and grade level in question, and $\bar{n}$ is the average

number of students per school. The SEM for a given performance category and grade level would be used to compute regions for the assignment of performance indices. In particular, if the quantity

$$\frac{\hat{O}_{cgs} - O_{cgs}}{SEM_{cg}}$$

is less than -1.5, $PI_{cgs} = 1$

between -1.5 and -.5, $PI_{cgs} = 2$

between - 5 and .5, $PI_{cgs} = 3$

between .5 and 1.5, $PI_{cgs} = 4$

greater than 1.5, $PI_{cgs} = 5$

where PI_{cgs} is the performance index in category c and grade level g at school s .